Europäisches Patentamt
European Patent Office
Office européen des brevets

Bescheinigung Certificate Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

02078418.7

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

Anmeldung Nr:
Application no.: 02078418.7
Demande no:

Anmeldetag:
Date of filing: 19.08.02
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

Koninklijke Philips Electronics N.V.
Groenewoudseweg 1
5621 BA Eindhoven
PAYS-BAS

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se referer à la description.)

Detecting and identifying anomalies in a signal

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s) revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

G11B20/00

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR IE IT LI LU MC NL PT SE SK TR

Detecting and identifying anomalies in a signal

1 Introduction

1.1 Playability

Besides appearance, functionality and prize, an important motivation to buy a certain drive is its ability to play back all sorts of discs without any problems. In the highly competitive market of optical storage systems, the ability of a drive to play back more different discs can get a drive producer an edge on its competitors. In this context the term *playability* can be defined as follows.

> *Playability is the ability of a non-ideal optical disc system to play back a disc without noticeable errors at the user side.*

This implies that a data system should deliver error free information and that an audio/video system must be allowed to perform error concealment to mask errors that cannot be prevented or removed.

1.2 Disturbances in optical disc systems

We can identify one particular group of disturbances for an optical disc drive that has to do with the quality of the optical disc. This quality can severely deteriorate due to incorrect or incautious handling of the discs by the user or the quality is bad from the start when the discs are poorly produced. One can think of scratches, dirt spots and fingerprints that arise on the polycarbonate substrate or the anomalies and impurities that are included in the substrate layer. These latter become of increasing importance again with the fast growing number of piracy discs entering the market.

From now on, we will call the disc related features we mentioned above *disc defects*. These defects, that are locally present on a disc, will distort the reflection of the laser beam. Hence they result in abnormal photoelectric signals that in turn will affect the generation of HF and servo signals and the behavior of all drive elements relying on these signals. See also Figure 1. Hence we treat these disc defects as disturbances according to the formal definition given above. The HF signal is further influenced by the geometry of the impressed pits and the sequence in which they appear on the disc. Anomalies in this pit/land structure are of a different origin and hence they will be excluded from the group of disturbances called disc defects. The following definition of the term disc defect will be used throughout this report.

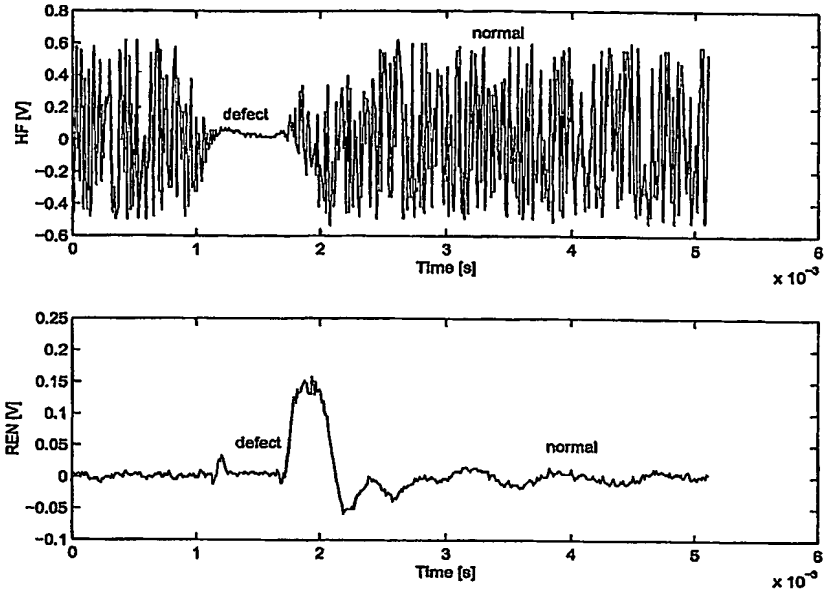


Figure 1: Example of a high-frequency signal (upper part) and a radial error signal (lower part), influenced by a disc defect.

> *Disc defects are those features locally present on or in an optical disc that result in different behavior of servo signals than what can be expected from the geometry of the information track and the dimensions or shape of the disc.*

Note that phenomena such as eccentricity, tilt and skew are excluded by this definition. This because these features are directly related to the track geometry or the disc dimensions and shape. They can also be caused by for instance the clamping of the disc inside the drive.

1.3 Dealing with disc defects

As already mentioned earlier there are methods to deal with disturbances that are influencing a system. An optical disc drive is equipped with several servo controllers that must assure the correct positioning of the laser spot on the information track. The way in which the servo tries to achieve accurate tracking is by constantly adjusting the laser position through an actuator in order to keep the positioning error equal to zero. The control algorithm determines how the actuator should be driven, based on the momentary error. During the design of these controllers the specifications for tracking performance and playability with respect to disc defects inevitably lead to a trade-off. Namely, for accurate tracking we want the controller to respond strongly to large position errors, which can be achieved by using a high bandwidth controller. Disc defects also result in, sometimes large, position errors. Since these errors are unreliable, ideally the controller should not respond to them at all, which implies a low bandwidth

controller. Even the use of more sophisticated controllers cannot improve the playability with respect to disc defects enough without sacrificing tracking performance. This is simply due to the fact that as soon as a disc defect occurs the photo-electric signals become severely distorted. These initial distortions can already influence the system in such a way that it stops functioning properly. More general we can state that for certain controlled systems disturbances exist that endanger the proper functioning of the system due to the severe response to these disturbances at their onset.

Therefore an accurate and fast detection mechanism is needed that can initiate proper actions in time and so preventing disturbances from influencing the system in a way that makes recovery impossible. A detection mechanism that is currently used monitors the total amount of reflected laser light. Whenever the reflected light intensity drops rapidly below a certain amount, a defect is detected. As a response the position error signals are artificially set to zero to prevent the controller from responding to false information. This detection however is not fast enough to prevent false position errors from influencing the controller at the start of the defect. This can already result in severe actuator drift that makes it hard to restart and continue tracking at the correct position when the defect is passed. Hence, in order to improve playability with respect to disc defects, next to improved position controllers also faster mechanisms for detection are needed in order to cross those disc defects.

Next to timely information on the occurrence of a disturbance it is also valuable to know what type of disturbance is entering the system. With this information available it becomes possible to select those countermeasures that yield the best results for a particular disturbance. Often however no adequate disturbance modelling techniques are available and the number of possible disturbances is infinitely large. In those cases the identification of disturbances can be based on a limited set of disturbance classes with whom new disturbances can be compared. Based on the outcome of this comparison, estimates can be made on the type of disturbance and its corresponding properties, that can be used in selecting a proper strategy to counteract its influences. When these disturbance classes are available they also can be used to design more specific and hence more accurate detection mechanisms for disturbance classes. This will further enhance the playability of optical disc drives.

1.4 Objectives

A precondition for the feasibility of 'design for playability' is the availability of fundamental knowledge and understanding of the limiting factors, possible interrelations and ways to deal with them. In this report the focus will be on several aspects of one of the influencing factors, namely disc defects as stated in the following problem definition.

> *How can the influence of disc defects be characterized and utilized on the basic engine level in order to achieve playability improvement?*

More precisely, the objectives of this research are the following.

- Development of a classification for disc defects based on available signals.
- Initiate improvements of defect detection.

However, this research will also concentrate on the implementation aspects of classification and detection methods in order to complete the outlines for further development. Although the research is conducted from the perspective of optical disc drives the objectives can be placed in the more general context of disturbances in data signals and controlled systems. Therefore the results will be generalized without loosing focus on optical disc drives where possible.

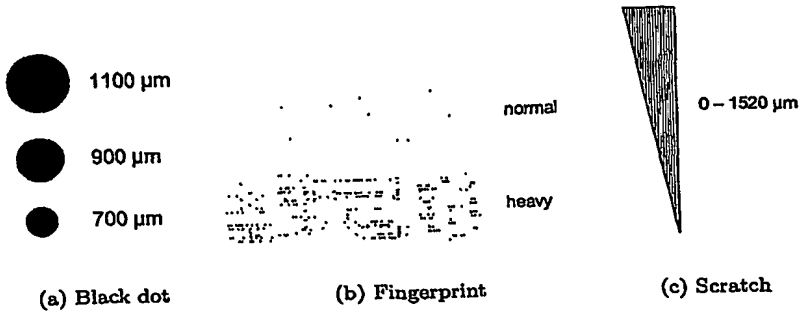


Figure 2: Artificial disc defects on standard test discs.

2 Defect measurements

An essential step in disc defect classification and detection is to understand how disc defects influence the various available signals in an optical disc drive. We proceed pragmatically by doing experiments and analyzing the results in order to gain these insights. Next to increased insight these experiments will also provide a vast amount of data that can easily be used in the research on disc defect classification and detection.

In Section 2.1 we deal with the issue of selecting representative disc defects and signals suitable for our experiments. In Section 2.2 we briefly address various practical issues concerning disc defect measurements.

2.1 Disc defect signals

2.1.1 Selection of defects and signals

In order to limit the number of experiments and the amount of data that will be generated, we carefully select a representative set of disc defects and signals to monitor during the experiments. This selection of a limited number of different disc defects is necessary since it is practically impossible to cover all different defects possibly present on an optical disc. For example one can easily make a dozen of scratches on a disc and none of them would have *exactly* the same shape. As a start we will use the various standardized defects that are used in optical disc drive development for testing purposes. Although the different ways in which the drive generates and processes signals for CD and DVD discs we limit ourselves to DVD test discs. Not only are the results of this research of greater importance for the more sensitive DVD systems but conclusions can relatively easily be extrapolated to the CD case.

The standard disc defects we will investigate are *black dots* of different sizes, normal and heavy *fingerprints* and artificial *scratches*. The first two artificial disc defects are produced by printing a single dot or a pattern of small dots on the substrate surface. For a heavy fingerprint the printed mesh is finer than for a normal fingerprint. This simulates a greasier, more spread out fingerprint on the substrate surface. The scratch is made by deliberately damaging a triangular area of the substrate surface with an abrasive tool. See also Figure 2.

Another standard 'disc defect' is the so-called *wedge*. This defect is used to simulate a damaged information layer, for instance caused by a scratch on the very thin protective layer at the disc's label side. However this defect is generated by replacing the normal data pattern with a random pit-land pattern. This electronic representation of a bad information track falls outside the stated definition of disc defects and hence wedges will not be taken into account.

With the above mentioned set of standardized disc defects already quit some different experiments can be conducted by varying the size of the defect and measuring on different locations inside the defect. For example we can measure signals when the outer edge of a black dot just falls within the laser beam and when the dot falls completely inside or even over the laser beam. In addition we extended the set of disc defects with some realistic ones that are made by deliberately abusing some new discs. These discs, now containing various radial and tangential scratches, dirt spots and fingerprints make the set of defects more representable and they can help to assess the validity of the artificial defects that only emulate reality. Another realistic but quit uncommon disc defect is a so-called *white dot*. At present no test discs are available for these defects that show a higher light reflection when compared to their surroundings. By placing some dots on a blanc DVD+RW disc before writing data to that area and removing the dots after the writing, these white dots can be simulated easily.

The photodetector of the OPU generates a number of photocurrents, depending on the exact configuration of the detector. Ideally we would measure these currents directly since this would rule out the influence of all signal processing steps and hence give us complete control on the way in which we generate the signals of interest from these currents. For DVD discs these currents however are not directly available on the test print and since these currents are in the order of several micro amperes in magnitude, measuring them directly on the engine PCB is not feasible. Signals that are available on the test print are the various servo signals such as the normalized radial and focus error (REN and FEN respectively), the normalized mirror (MIRN) signal that is a measure for the total amount of laser light received by the photodetector, the normalized tilt signal and of course the HF signals and various derivatives used in the engine decoder and data path. From preliminary experiments and previous research it became clear that both the MIRN and HF signal behavior show the most direct relation with incoming disc defects. We note however that the latter contains a high frequency component that carries the digital data. This component can be regarded as noise when investigating disc defect influences that occur in a lower frequency range. Next to the MIRN signal we also want to monitor the behavior of the REN and FEN signals since these signals are directly involved in the positioning of the laser spot and reducing the influence of disc defects on this positioning is precisely what we are aiming at. It is clear that the REN and FEN signals are not reliable during the occurrence of a disc defect. The fact that the laser spot position is adjusted by a closed-loop control system increases this uncertainty. Hence care must be taken when analyzing measurements of these signals.

2.2 Measuring defect signals

The signal measurements are conducted at a sampling frequency f_s of 500 kHz and for each signal 512 samples are taken. This results in a total measuring

time of 1.024 *ms*. The sample frequency is fixed for all measurements to make comparison of measurements possible. Although it is possible to elongate the measuring time by taking in more samples, all experiments are conducted with the same acquisition length to limit the amount of data and required processing time. In order to start the measurements, we use the output signal of the currently implemented defect detector (DEFO) as trigger signal. By including a certain amount of 'pre-trigger' data in the time series, the slow response of the detector will not lead to any loss of information. In cases where no accurate triggering is possible with the DEFO signal, we apply a trigger based on the MIRN signal.

All measurement results, from which we will present only a small portion, are retrieved from the oscilloscope and stored in Matlab data files, since this program provides extensive capabilities for numerical data processing. By using the 'structure' data format not only the time series can be stored but in combination also various measurement settings and additional comments on the experiments are saved. This makes it easy to identify particular measurements when the number of experiments grows. The data for each measurement is stored in separate files that are named following a strict convention. By using this convention it is possible to automate the process of data retrieval for a large number of experiments. The benefits of this approach become even more clear in the next chapter.

3 Defect classification

Timely knowledge of the type of disc defect that is influencing the optical disc drive can help to improve servo performance and hence playability. This information makes it possible to select or adjust control strategies and other countermeasures to eliminate influences of disc defects on the system. Since parametric models of signals affected by disc defects are not yet available, estimation methods like for instance a Kalman filter, cannot be used to identify disc defects.

Identification of disc defects by comparing new signals with a database of known defect signals resolves this problem as long as the database contains enough measurements. Given the enormous number of possible disc defects the feasibility of this method is limited by the available memory for the database and the speed of algorithms to search through the stored data. The size of a database with reference signals can be reduced by identifying a limited number of classes that each describe a large group of defect signals in the whole data set. These defect classes can also be used to design more specific defect detection strategies. Such detectors only have to detect particular groups of disc defects instead of having one detector looking for the occurrence of all possible defects. The whole disc defect classification process is schematically presented in Figure 3.

3.1 Hierarchical clustering

An important clustering method is the hierarchy. This structure will form the basis of the disc defect classification that we treat in more detail in Section 3.2. In general a hierarchy is a set $\mathcal{S}_{\mathcal{H}} = \{\mathcal{S}_h : h \in \mathcal{H}\}$ of subsets $\mathcal{S}_h \subseteq \mathcal{I}$, $h \in \mathcal{H}$, called *clusters* and satisfying the following conditions:

1. $\mathcal{I} \in \mathcal{S}_{\mathcal{H}}$;
2. for any $\mathcal{S}_1, \mathcal{S}_2 \in \mathcal{S}_{\mathcal{H}}$, either they are non-overlapping ($\mathcal{S}_1 \cap \mathcal{S}_2 = \emptyset$) or one of them includes the other (divisive: $\mathcal{S}_1 \subseteq \mathcal{S}_2$ or agglomerative: $\mathcal{S}_2 \subseteq \mathcal{S}_1$), all of which can be expressed as $\mathcal{S}_1 \cap \mathcal{S}_2 \in \{\emptyset, \mathcal{S}_1, \mathcal{S}_2\}$;
3. for each $i \in \mathcal{I}$, the corresponding singleton is a cluster, $\{i\} \in \mathcal{S}_{\mathcal{H}}$.

3.2 Disc defect clustering

As mentioned we want to classify disc defects and use the results to identify new disc defects by comparing them with the known reference defects derived from the classification. This approach relaxes the need for accurate physical models of disc defect and their influence on measurable signals. In addition it also drastically reduces the amount of data that is needed for defect identification. Finally the approach of classification also provides us with the required selection criteria to identify defects.

We intend to use measurements of servo signals that are affected by various disc defects for this purpose. Therefore, according to the definitions in the previous paragraph, we actually need to develop a clustering algorithm. From the resulting arrangement of disc defect measurements in distinct groups, we then can derive a formal disc defect classification. Note that these classes themselves are part of a larger classification that covers all possible disturbances in optical disc drives.

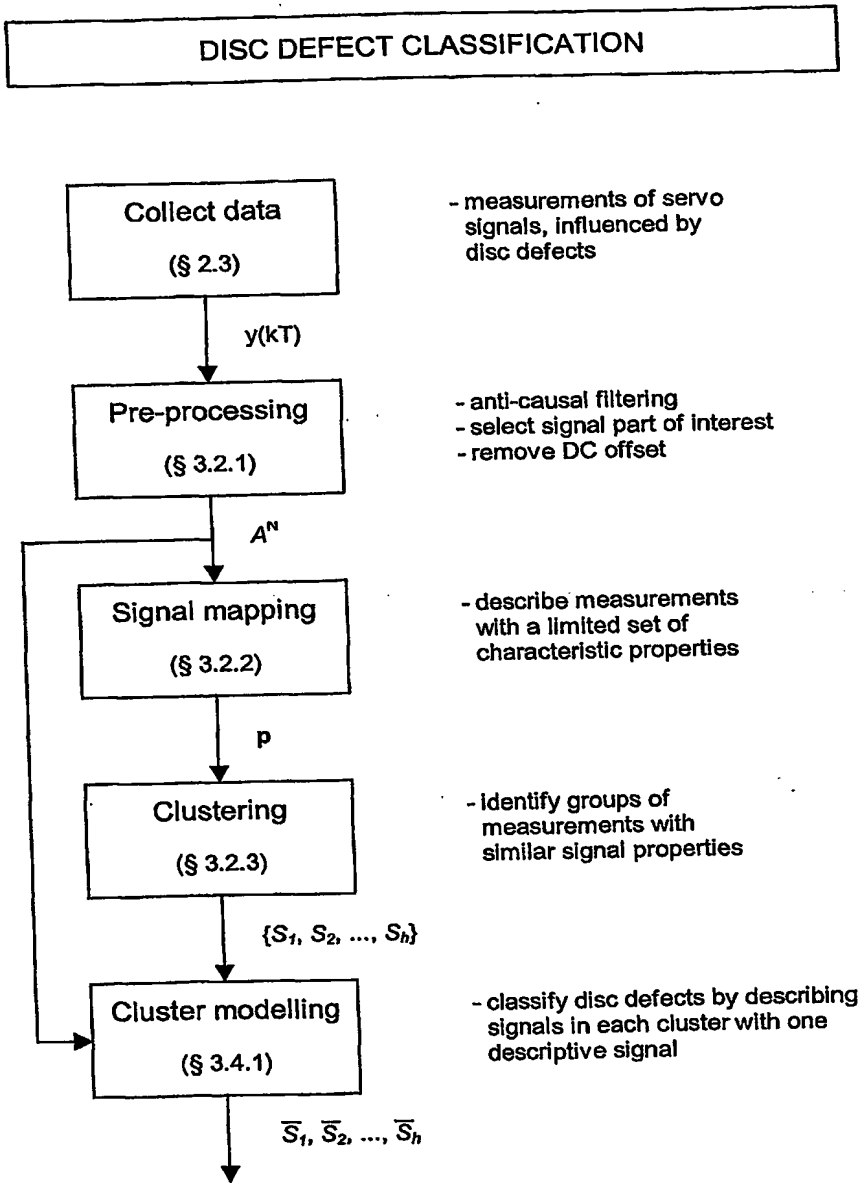


Figure 3: Disc defect classification flowchart.

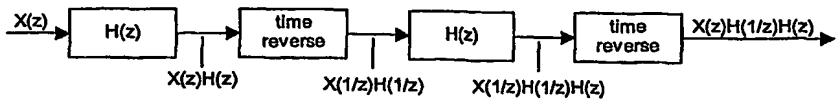


Figure 4: Anti-causal, zero-phase filtering process.

3.2.1 Signal processing

In the subsequent analysis we primarily focus on measurements of the MIRN signal. The physical understanding on how this signal is influenced by disc defects helps in the development and the validation of a clustering algorithm. The unreliability of the REN and FEN signals and the lack of physical understanding of their behavior during disc defects, further justifies the focus on the MIRN signal. In Section 3.3 we briefly come back to this subject.

We expect the MIRN signal to be locally 'constant' when there are no severe disturbances such as shocks or disc defects present. This assumption can be confirmed by examining the measurement results presented in Chapter 2. Those graphs however also show that the measurements are corrupted with noise. It is assumed that the observed noise is a combination of measurement noise, internally generated system noise and quantization noise caused by the analogue-to-digital conversion in the measuring device. In order to remove this noise, which can obscure the phenomena we are interested in, we start by filtering the measurements.

For the filtering of the time series we use an anti-causal, zero-phase digital filter implementation. The benefit of this type of filtering is that it makes it possible to eliminate all phase shifts introduced by the filtering. By further choosing a gain at low frequencies equal to one, the exact shape of the signal is preserved. The general processing scheme for this filter implementation is depicted in Figure 4. The reversing of the filtered time series and filtering this sequence again removes all phase shifts introduced in the first filter passage. Note that the time reversing of the filtered sequence also introduces the non-causality mentioned. Hence this type of zero-phase filtering is only possible with time series that are completely available off line. Hence this method to remove measurement noise can not be used in an online clustering algorithm. Figure 5 shows the result of the whole filtering process on a particular defect measurement.

From the measurement results presented in Chapter 2 we also note that, for most disc defects, the largest part of the time series is unaffected. Since we are only interested in the part of the MIRN signal that deviates from its normal 'constant' level, these constant parts of the measurement are removed. This is done by selecting the region of interest for each measurement by hand. Not enough information is available to exactly define the begin and end of the affected regions automatically. However with some physical insight, objective inspection of the data and common sense the selection can be made relatively accurate as shown in Figure 6. This division between the normal and disturbed MIRN behavior also offers us the possibility to determine the DC offset of the measured signals and remove it from the time series. This is done by determining the average value of the normal signal regions and subtracting this offset value from the complete time series.

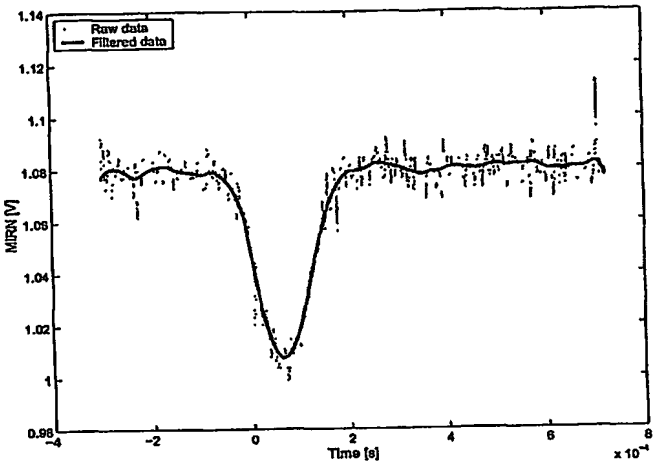


Figure 5: Reducing noise and preserve shape with zero-phase filtering.

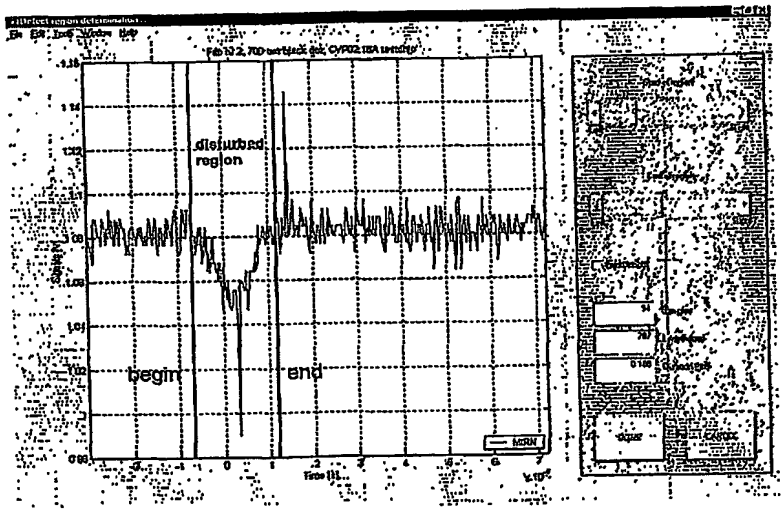


Figure 6: Selecting begin and end of the affected region in a measurement.

3.2.2 Describing defect signals

From literature it appears that the clustering method depends on the kind of input data that is available. In our case we want to cluster a set of time series. However we believe that it is not wise to use these time series directly. First of all the length of the time series depends heavily on the type of disc defect and the exact location of the laser spot on the disc. Moreover the measurements are still, despite the filtering, corrupted with noise. Therefore we need to describe the signal with a set of characteristic parameters that are insensitive to those variations. Preferably this set is chosen small in order to limit the computational efforts.

A good set of characteristic variables would be the parameters of a model that describes the disc defect signal. We already pointed out that no suitable parametric models are available yet. Using a black-box identification method could resolve this problem. However, to choose the right identification method in order to get a set of robust parameters, still quit some insight and knowledge of the system is required. Another possibility is to describe the signals with a set of statistical quantities. The non-stationary[1] character of the signals under study complicates this approach. More importantly, the assumption of randomness interferes with our previous statement that disc defect signals in essence are of a deterministic nature.

To overcome these problems we approach the problem in a more intuitive manner. By combining various concepts from the theories discussed above with the insights gained on disc defect signals, a suitable set of characteristic signal properties is found. Note that from now on we assume that the DC offset of the measured MIRN signals is removed as we discussed in Section 3.2.1. This implies that a MIRN level around zero corresponds to a normal light reflection and higher or lower MIRN values indicate an increased or decreased reflection relative to the normal situation. Furthermore we only consider the affected region that we previously selected by hand from the whole time series.

The first characteristic property is the *mean value* of the disc defect signal. This value is particularly useful to distinguish between defects that have a higher and lower reflectivity than the normal disc. For a black dot with a lower reflectivity than the normal disc, the mean value will be below zero while it will be positive for a white dot. The second characteristic parameter is the *duration* that, more conveniently, can be expressed through the number of measured samples when the sample time of the measuring device is known. The third property is the *peak value* of the disc defect signal. To make a fair comparison between values for all disc defects, the absolute peak value is taken. Otherwise the peak value for a white dot would always be higher than it would be for a black dot. This is undesirable since we are mainly interested in the different behavior of signals compared to the normal situation. Finally we divide the signal into a fixed number of *amplitude bands* and count the number of samples that fall within each band. See also Figure 7. The resulting values for each amplitude band complete our set of characteristic parameters. It is not likely that all these signal properties yield a value in the same order of magnitude. Therefore we add weighting factors to all parameters in order to obtain a balanced set of signal properties.

[1] A random process whose statistical properties are invariant in time is called stationary. Clearly this is not the case for the defect signals under consideration.

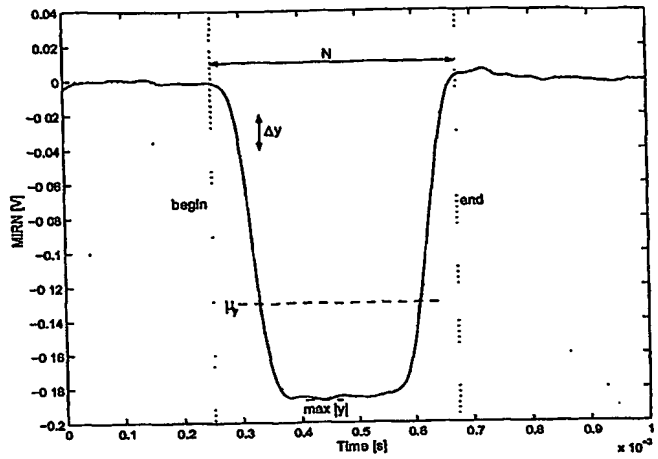


Figure 7: Properties used to describe the signals of interest.

Formally we can summarize the results of this section as follows. Let $\mathcal{I}$ denote the set of all possible time series for the MIRN signal. The subset $\mathcal{A}^N \subset \mathcal{I}$ is formed by L time series of length N, representing all MIRN signals that are affected by disc defects and disc defects only. This subset contains all N-dimensional vectors $\mathbf{y}_r = (y_{r1}, y_{r2}, \ldots, y_{rN})$, representing the affected MIRN time series in $\mathcal{I}$. The index r can be seen as a label that uniquely identifies the corresponding element $i \in \mathcal{I}$. Next $p_1, p_2, \ldots, p_m$ span a space $\mathcal{B}$ in $\mathbb{R}^m$, where p_j, $j = 1, 2, \ldots, m$ are the m different descriptive properties for the elements in $\mathcal{I}$. We can now define the mapping $\mathcal{F}$: $\mathcal{A}^N \subset \mathcal{I} \rightarrow \mathcal{B}$ that maps the time series of interest (domain) to a set of descriptive signal properties (range). This mapping can be further specified by defining the functions f_j : $\mathcal{A}^N \subset \mathcal{I} \rightarrow p_j$, $j = 1, 2, \ldots, m$. From the selected set of characteristics we discussed earlier, these functions are defined as follows.

$$f_1(y_r) = W_1 \cdot \overline{y}_r = W_1 \cdot \frac{1}{N}\sum_{k=1}^{N} y_r(k) \tag{1}$$

$$f_2(y_r) = W_2 \cdot N \tag{2}$$

$$f_3(y_r) = W_3 \cdot \max|y_r(k)|, \quad k = 1, 2, \ldots, N \tag{3}$$

$$f_4(y_r) = W_4 \cdot \sum_{k=1}^{N} \begin{cases} 1, & (m-4)\Delta y \leq |y_r(k)| < (m-3)\Delta y \\ 0, & (m-4)\Delta y > |y_r(k)| \vee (m-3)\Delta y \leq |y_r(k)| \end{cases} \tag{4}$$

$$\vdots$$

$$f_m(y_r) = W_m \cdot \sum_{k=1}^{N} \begin{cases} 1, & (m-4)\Delta y \leq |y_r(k)| \\ 0, & (m-4)\Delta y > |y_r(k)| \end{cases} \tag{5}$$

where k is the running variable representing the sample instant, N is the total number of samples in the affected region of the signal and the index r indicates on which element of $\mathcal{A}^N \subset \mathcal{I}$ the mapping is performed. The number of descriptive properties m is defined by the number of amplitude bands that is used

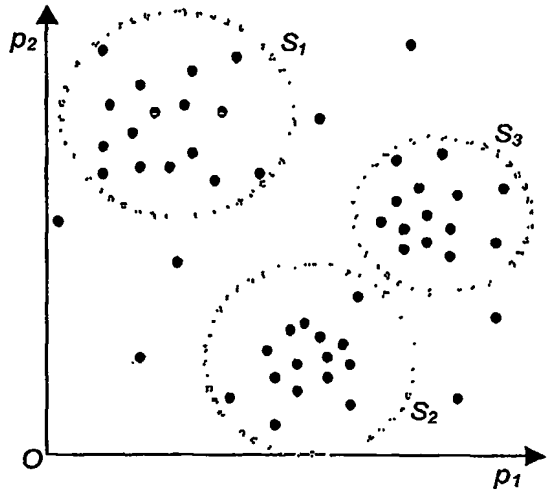


Figure 8: Geometric interpretation of signal mapping and clustering for $n = 2$.

in the discrete amplitude distribution. See (4) and (5). $W_1, W_2, \ldots, W_m$ are the weighting factors that can be used to balance the signal mapping.

3.2.3 Clustering algorithm

With the mappings presented in (1) to (5) L different m-dimensional property vectors $\mathbf{p}_r = (f_1(y_r), f_2(y_r), \cdots, f_m(y_r))$, $r = 1, 2, \ldots, L$ can be constructed. These row vectors can also be interpreted as unique points in $\mathcal{B}$, representing the corresponding MIRN time series in $\mathbb{R}^m$. In general, clustering is the process of distinguishing separate groups of similar data. In analogy to the geometric interpretation of the MIRN signal mapping, clustering can be seen as the identification of different groups of closely spaced data points. This is depicted graphically in Figure 8 for the two-dimensional case. For the various clusters ($\mathcal{S}_1$, $\mathcal{S}_2$ and $\mathcal{S}_3$ in Figure 8) it holds that the geometrical distances between points within one cluster are much smaller than the distances between points belonging to separate clusters.

A clustering method that directly uses this geometric interpretation of similarity is agglomerative hierarchical clustering. The input for this clustering method is a so-called dissimilarity entity-to-entity matrix, where each entity is considered as a single cluster or singleton, denoted by $\mathcal{S}_h$, $h \in \mathcal{H}$. Note that $\mathcal{H}$ is the set of all cluster labels and that each h is uniquely related to one cluster. For an agglomerative hierarchical clustering of L objects, the set $\mathcal{H}$ holds $2L - 1$ labels, where the first L elements correspond to the original entities or singletons. The dissimilarity matrix can easily be derived from the mapped data points by calculating the distance between every pair of objects in the data set. From literature various definitions for vector distances are available.

Euclidean distance

$$d_{rs} = \sqrt{(\mathbf{p}_s - \mathbf{p}_r)(\mathbf{p}_s - \mathbf{p}_r)^T} \tag{6}$$

City Block distance

$$d_{rs} = \sum_{j=1}^{m} |p_{rj} - p_{sj}| \tag{7}$$

Minkowski metric

$$d_{rs} = \left\{ \sum_{j=1}^{m} |p_{rj} - p_{sj}|^p \right\}^{1/p} \tag{8}$$

A more general notation. When $p = 1$ it represents the City Block distance, and when $p = 2$, this metric is equal to the the Euclidean distance.

The indices r and s denote the labels for the corresponding clusters. The most widely used distance measure is the Euclidian distance, usually denoted as $\|\mathbf{p}_s - \mathbf{p}_r\|$. Also note the similarity of the above distance measures with the concept of vector norms. With the above distance measures a dissimilarity matrix $D = [d_{rs}]$ with $r, s = 1, 2, \ldots, L$, can be constructed. Note that D is symmetric and the elements of its main diagonal are zero. With the dissimilarity matrix available the main steps of the algorithm are as follows.

Step 1 Find the minimal value $d(r^*, s^*)$, $r^* \neq s^*$ in the dissimilarity matrix, and form the merged cluster $\mathcal{S}_h = \mathcal{S}_{r^*} \cup \mathcal{S}_{s^*}$, $h \in \mathcal{H}$.

Step 2 Transform the dissimilarity matrix by substituting one new row (and column) h for the rows and columns r^*, s^*, with its dissimilarities defined as

$$d(r, s) = F(\{\mathcal{S}_r\}, \{\mathcal{S}_s\}, l_r, l_s) \tag{9}$$

with $r, s \in \{1, 2, \ldots, h\} \cap \{r^*, s^*\}'$. F is a fixed dissimilarity function and l_r, l_s define the number of objects in cluster $\mathcal{S}_r$ and $\mathcal{S}_s$ respectively. If the number of clusters obtained is larger than 2, go to Step 1, else End.

The function F defines the dissimilarity between the merged clusters. Since these clusters can contain more than one object, the distance measures, as defined in (6) to (8), cannot be used here. Several popular methods to define the inter-cluster distance or dissimilarity are presented below.

Nearest neighbor (Single linkage) uses the smallest distance between objects in the two clusters $\mathcal{S}_r$ and $\mathcal{S}_s$.

$$d(r, s) = \min \|\mathbf{p}_{sj} - \mathbf{p}_{ri}\|, \quad i \in (1, \ldots, l_r), \; j \in (1, \ldots, l_s) \tag{10}$$

Farthest neighbor (Complete linkage) uses the largest distance between objects in the two clusters.

$$d(r, s) = \max \|\mathbf{p}_{sj} - \mathbf{p}_{ri}\|, \quad i \in (1, \ldots, l_r), \; j \in (1, \ldots, l_s) \tag{11}$$

Average linkage uses the average distance between all pairs of objects in the two clusters $\mathcal{S}_r$ and $\mathcal{S}_s$.

$$d(r, s) = \frac{1}{l_r l_s} \sum_{i=1}^{l_r} \sum_{j=1}^{l_s} \|\mathbf{p}_{sj} - \mathbf{p}_{ri}\| \tag{12}$$

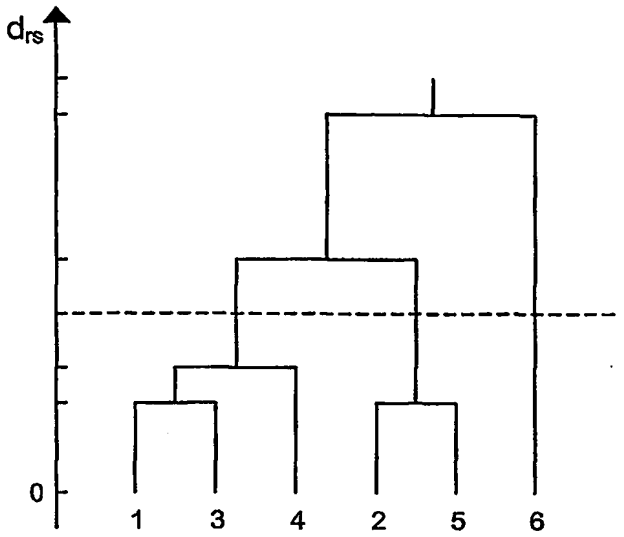


Figure 9: Graphical representation of a hierarchical cluster tree.

Centroid linkage uses the distance between the centroids of the two groups $\mathcal{S}_r$ and $\mathcal{S}_s$.

$$d(r, s) = \|\overline{\mathbf{p}}_s - \overline{\mathbf{p}}_r\| \tag{13}$$

where:

$$\overline{\mathbf{p}}_r = \frac{1}{l_r} \sum_{i=1}^{l_r} \mathbf{p}_{ri} \tag{14}$$

and $\overline{\mathbf{p}}_s$ is defined similarly.

Ward linkage uses the incremental sum of squares; that is, the increase in the total within-group sum of squares as a result of merging clusters $\mathcal{S}_r$ and $\mathcal{S}_s$.

$$d(r, s) = l_r l_s \frac{d_c^2(r, s)}{l_r + l_s} \tag{15}$$

where $d_c^2(r, s)$ is the squared distance between clusters $\mathcal{S}_r$ and $\mathcal{S}_s$ defined in the Centroid linkage by (13).

The results of the agglomerative hierarchical clustering method can be represented graphically as a tree. An example of such a hierarchical cluster tree or *dendrogram* is shown in Figure 9. In such a graph the numbers at the horizontal axis represent the indices of the original singletons and they are called *leaf nodes*. The links between the objects are represented by the connecting horizontal lines, called interior nodes. The height of the vertical link lines indicate the distance between the linked objects. Note that disproportionately long vertical lines can indicate that the corresponding objects are combined incorrectly. With this graphical representation of the cluster tree we can easily define an arbitrary number of clusters C by drawing a horizontal line in the dendrogram. All the leaf nodes—representing the entities—that are connected below this line

belong to one particular cluster $c \in C$ with $C = \{1, 2, \ldots, C\}$. For the example shown in Figure 9, three clusters can be defined by drawing a dividing line such that it only bisects three vertical lines in the tree. This results in the clusters $\mathcal{S}_1 = \{1, 3, 4\}$, $\mathcal{S}_2 = \{2, 5\}$ and the singleton $\mathcal{S}_3 = 6$.

The presented linkage methods will all give the same or almost the same results, when applied to well-structured data. When the structure of the data is somewhat hidden or complicated, the methods may give quit different results. In the latter case the single and complete linkage methods represent the two extremes of the generally accepted requirement that the 'natural' clusters must be internally cohesive and, simultaneously, isolated from the other clusters. Single linkage clusters are isolated but can have a very complex chained and noncohesive shape. In contrast the complete linkage clusters are very cohesive, but may not be isolated at all. The other three methods result in a trade-off between cohesiveness and isolation of the resulting clusters.

3.3 Cluster results and validation

In this section we present the results of the disc defect clustering, obtained with the agglomerative, hierarchical algorithm. From the various options presented in Section 3.2.3 we select the Euclidean distance measure and Ward linkage method. The *Euclidean distance measure* is selected since it is easy to calculate and its geometrical interpretation is straightforward. The *Ward linkage method* is chosen since it provides a good trade-off between cluster cohesiveness and isolation. When compared to similar methods (average and centroid linkage) the Ward linkage appears to result in the most logical clustering based on analysis of the corresponding defect signals and their physical interpretation.

The results of the clustering process for the set of reference measurements is shown graphically in Figure 10. In the dendrogram we observe two inconsistent links, denoted by **A** and **B**. The first one links the artificial scratch of 1320 μm to, among others, the quarter black dots while we would expect them to be linked to the 1120 and 1420 μm artificial scratches. Figure 11 reveals the cause of this inconsistency. It appears that the MIRN signal has been normalized differently during the measurements with the 1320 μm scratch. After removing the DC offset and comparing different scratch measurements, the lowest MIRN level will be higher for the 1320 μm scratch. From this one could conclude that this particular scratch reflects more light than others, which is not true in reality. This MIRN level difference is however adequately translated by the signal mapping and hence this particular scratch is placed in another cluster. During new measurements the phenomenon kept reappearing so most likely the cause lies in a local variation of the reflection of the test disc. In our further analysis we will carefully check wether this phenomenon influences the clustering results.

The second inconsistency we observe in the dendrogram is related to the different fingerprints in our defect database. The clustering algorithm assigns a disproportionally large difference to the normal fingerprints and the heavy fingerprint, indicated by the relatively long vertical lines below **B**. Although the two types of fingerprints clearly differ, these differences are considered to be less significant than, for example, those between a small edge black dot and a large radial scratch. See also Figure 12. The cause of this inconsistency lies in the different amplitude drops of the MIRN signal for normal and heavy

Figure 10: Dendrogram of the disc defect clustering; Euclidean distance, Ward linkage, $W_m = 1$, $m = 1, 2, \ldots, 14$.

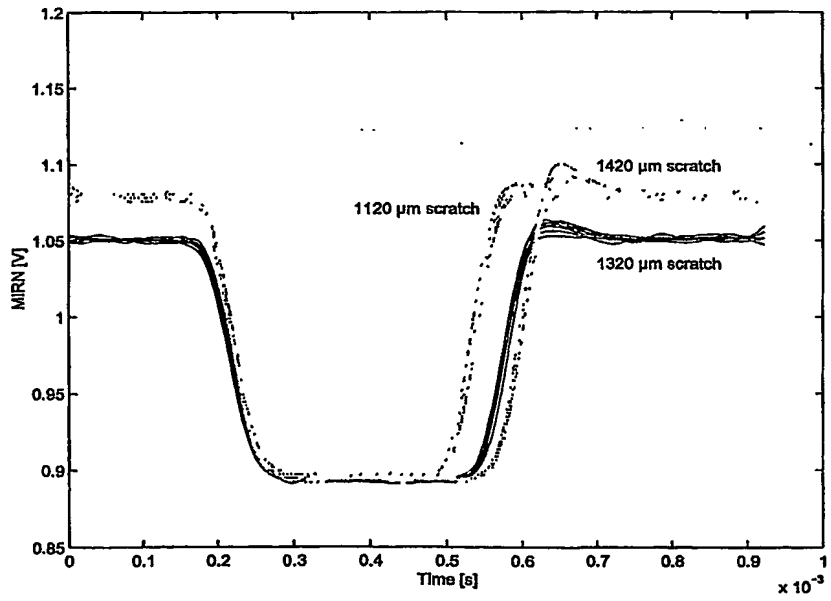
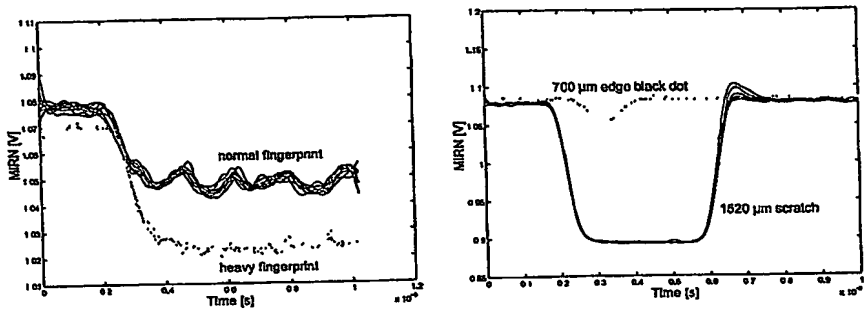


Figure 11: Different MIRN normalization for 1320 μm artificial scratch measurements.



(a) Normal and heavy fingerprints

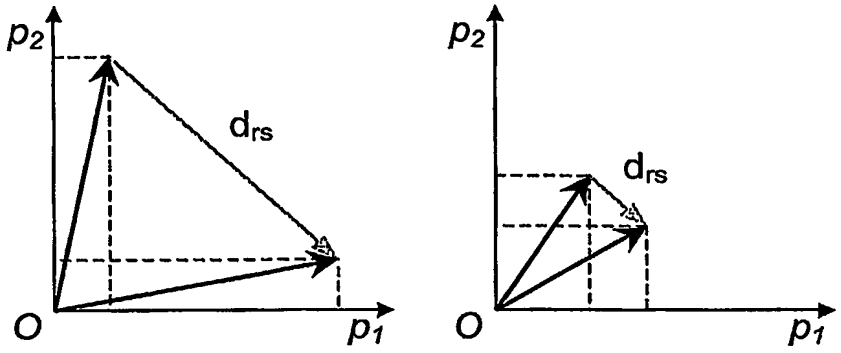
(b) Black dot edge and radial scratch

Figure 12: Comparing normal and heavy fingerprints, edge black dots and radial scratches.

fingerprints respectively. Due to the long duration common to fingerprints, a high number of samples will fall inside one particular amplitude band. The amplitude difference for normal and heavy fingerprints causes this high property value to appear at a different column in the property vector. The effect of this on the distance between objects is shown graphically in Figure 13. In order to reduce the mentioned inconsistency we must make the property vector more robust for small amplitude variations. This can be done by adjusting the weighting factor in (2). The value for the weighting factor is determined by trial and error. The resulting value $W_2 = 5$ yields a more balanced mapping and hence the clustering results are more in line with our interpretation of the defect signals.

Another clustering 'mistake' is the combination of black dots and white dots in one cluster. Based on visual inspection and physical interpretation of the corresponding defect signals, we would place these defects in separate clusters, since these two types of defects are more or less opposites. The obvious difference between the two types of defects is shown again in Figure 14. Although this phenomenon does not result in inconsistent links in the dendrogram it is closely related to the fingerprint case we discussed. The only signal property suitable for making a distinction between higher and lower reflection is the signal mean value from (1). The other properties are all based on the absolute MIRN signal and hence give similar results for both black dots and white dots as can be seen from Figure 14. By adjusting the weighting factor in (1), a better distinction between black dots and white dots is achieved. By trial and error the weighting factor in $W_1 \cdot f_1(y_r)$ is determined, yielding $W_1 = 1 \cdot 10^4$.

The clustering results from the adjusted algorithm are depicted graphically in Figure 15. Comparing this result with the previous dendrogram in Figure 10 shows that the added weighting factors indeed reduce the inconsistency significantly. Table 1 summarizes which disc defects are grouped together in the different clusters that are selected in the dendrogram. From this table it becomes clear that with the adjusted algorithm also the 1320 μm scratch is clustered according to our expectations. As we mentioned before we can easily alter the



(a) Sensitive vectors (3, 15) and (15, 3)

(b) Robust vectors (5, 8) and (8, 5)

Figure 13: Dissimilarities caused by unbalanced mapping.

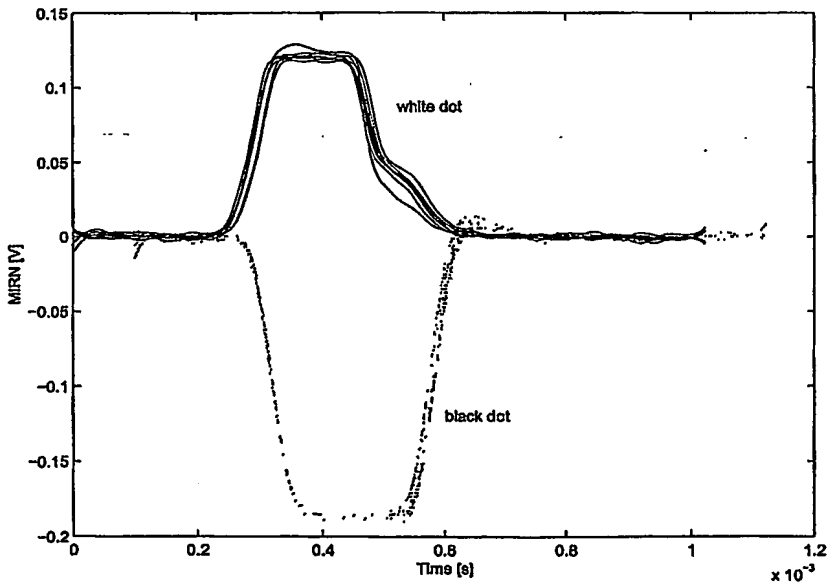


Figure 14: Different signal behavior for black dots and white dots.

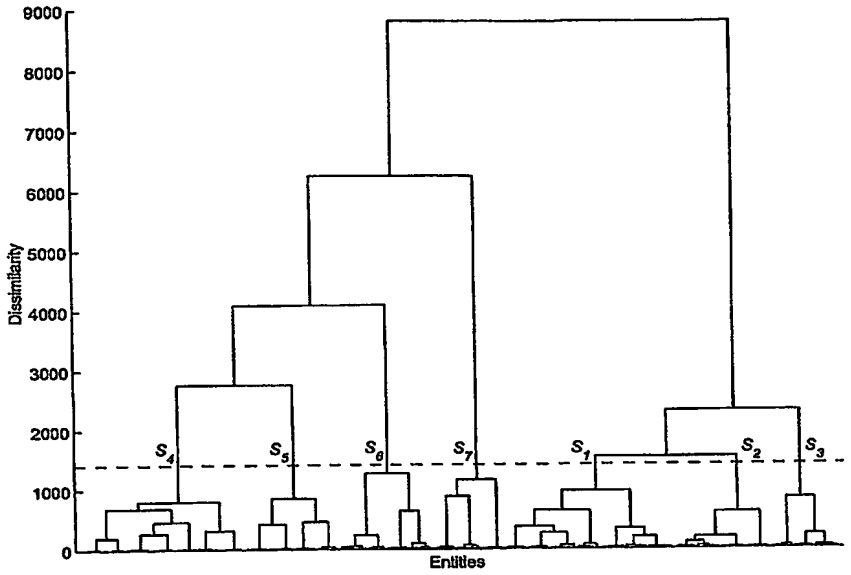


Figure 15: Consistent dendrogram of the disc defect clustering; seven clusters selected, Euclidean distance, Ward linkage, $W_1 = 1 \cdot 10^4$, $W_2 = 5$.

Table 1: Objects in disc defect clusters.

Cluster	Objects
$\mathcal{S}_1$	middle black dot 700 μm middle black dot 900 μm scratch 420–820 μm
$\mathcal{S}_2$	middle black dot 1100 μm scratch 920–1120 μm
$\mathcal{S}_3$	scratch 1320–1520 μm
$\mathcal{S}_4$	edge black dot 700–900 μm quarter black dot 700–900 μm scratch at R = 32 mm
$\mathcal{S}_5$	scratch 320 μm scratch at R = 32 mm scratch at R = 35 mm
$\mathcal{S}_6$	all white dots
$\mathcal{S}_7$	all fingerprints

Table 2: Combining disc defect clusters.

Number of clusters	Combined clusters
6	$\mathcal{S}_1 \cup \mathcal{S}_2$
5	$\mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3$
4	$\mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3$ $\mathcal{S}_4 \cup \mathcal{S}_5$
3	$\mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3$ $\mathcal{S}_4 \cup \mathcal{S}_5 \cup \mathcal{S}_6$
2	$\mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3$ $\mathcal{S}_4 \cup \mathcal{S}_5 \cup \mathcal{S}_6 \cup \mathcal{S}_7$

number of clusters by shifting the bisecting line up or down in the dendrogram. Table 2 shows which clusters are combined when a smaller number of clusters is selected. Both from Table 2 and the dendrogram in Figure 15 the existing hierarchy in the disc defect database becomes clear. Actually two major clusters can be identified. One with all middle black dots and artificial scratches and the other holding all the other defects. In both these large groups a further subdivision of disc defect types can be made, which is clearly depicted by the step-like linkage in the right and left parts of the dendrogram. Close examination of the various combinations of disc defects shows that this clustering is according to our expectations, based on physical interpretation of the different disc defects.

The measured MIRN signals, corresponding to the disc defects in each cluster are shown in Figure 16. Note that the correlation between all signals in the same cluster is optimized to prevent the differences in used pre-trigger time and defect duration from cluttering the general view. Figures 17 and 18 show the corresponding radial and focus error signals respectively. In these figures we observe that the REN and FEN signals also reveal some characteristic distinctions between the various clusters. The dissimilarities however are far less distinct than those for the MIRN signals. Applying the clustering algorithm to the REN and FEN signal therefore does not give satisfactory results. However the results make us believe that with a different or extended set of mapping functions, specifically chosen for the REN and/or FEN signals, better clustering results can be obtained with the REN and FEN signals. Another possibility that could lead to even better and distinctive disc defect clusters is to use a combination of the mappings for the MIRN, REN and FEN signals.

3.4 Cluster modelling

Now the disc defect clusters are available, one step of the classification process remains. The clusters themselves are nothing more than distinctive groups of similar defect signals. The data reduction we strive for is however not yet achieved. In this section we discuss how we can obtain an adequate description for each cluster that answers to this classification objective.

The most suitable class description is that of a signal or model from which such a signal can be derived. Not only is it easy to compare graphs of different

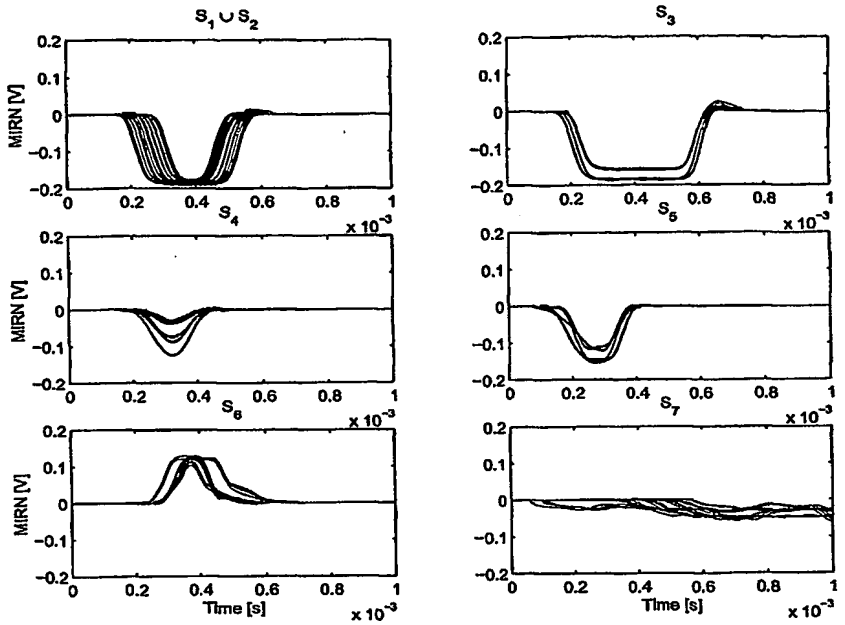


Figure 16: MIRN signals for clustered disc defects.

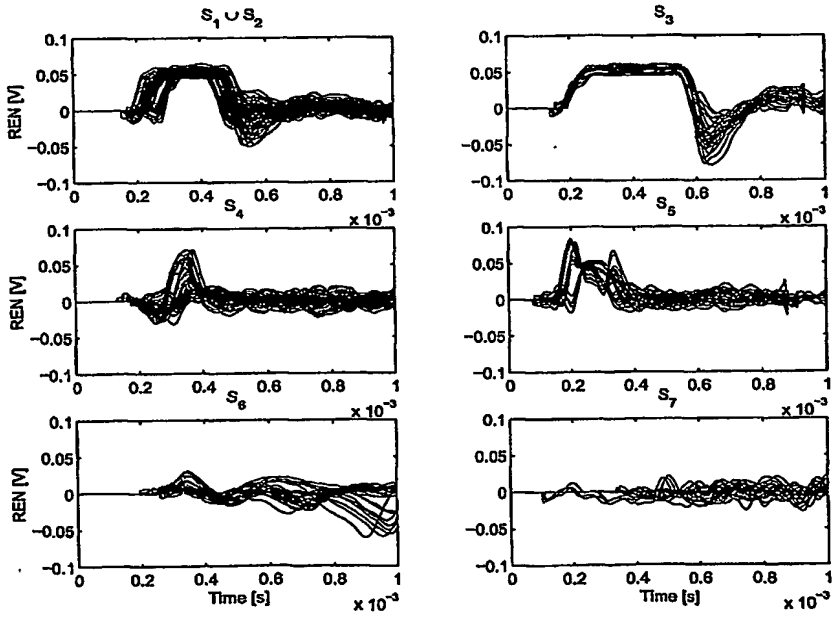


Figure 17: REN signals for clustered disc defects.

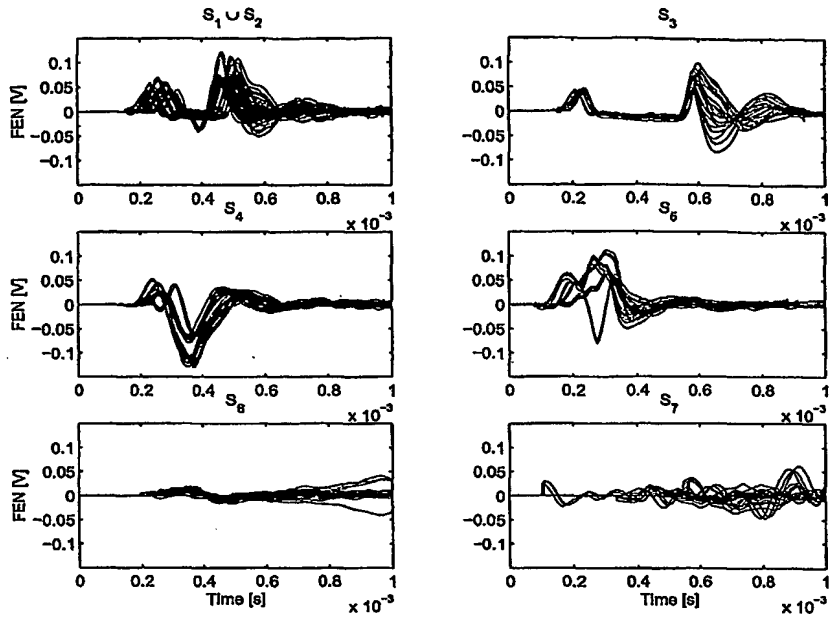


Figure 18: FEN signals for clustered disc defects.

signals qualitatively but a signal, or better a time series, can be used directly in mathematical computations. A class description in words for example, lacks this possibility. The problem now is to derive a representative signal (or model) that adequately describes all the signals belonging to one cluster. Inevitably a trade-off must be made between the accuracy of the description for individual signals and its general validity for the whole cluster.

A straightforward method for this task is to fit a function to the time series in the cluster (see Figure 16) that approximates the data according to some criterion. The key issues for this approach are the choice for a general form of the function and the selection of a suitable criterion. The most widely used criterion is the sum of the squares of the errors between the fitted function and the data points. Methods using this criterion are usually denoted as least squares (LS) methods.

Preferably the function or model structure is based on (physical) laws that relate the signals to the system that generates them. When such a structure is unavailable a more general structure must be used. Examples of such general function structures are the Fourier and Prony decomposition that approximate the data with a sum of sinusoidal or complex exponential functions respectively. Other possibilities are to approximate the data with polynomials or splines.

3.4.1 Least squares polynomial fitting

To obtain descriptive signals for each disc defect cluster we will apply a least squares polynomial fitting method. We choose a polynomial to approximate the defect signals in the first place because we lack a parametric function structure

that is based on a disc defect model. Fourier and Prony decomposition are not usable since the disc defect signals hardly show any periodic behavior, except for the small oscillations we observed in the MIRN signals during the passage of a fingerprint. A good alternative would be to use splines for the approximation when a spline fitting algorithm is available that can deal with several signals simultaneously.

The results of the fitting with a polynomial of degree $n = 15$ are shown in Figure 19. From this figure we conclude that the time series obtained from the fitted polynomial functions describe the disc defect classes reasonably well. However, due to the nature of the fitted function, we observe some small oscillations in the resulting signal that are not present in the original time series. Especially at the edges of the defect signal these deviations can become significant when we tend to use the defect class signals for detection. For that purpose the begin and end regions of the signal must be known as accurately as possible. Applying a fitting routine that uses splines could resolve this since splines offer the possibility to impose demands on the slope of the fitted signal in regions where additional accuracy is desired.

3.4.2 Class validation

Now the classification of our set of reference disc defects is complete we can perform some tests in order to validate the method. This is done by applying the property mapping from Section 3.2.2 to the derived class signals and do the same for some well-known, new defect measurements. Then we calculate the distance between the property vector of the test measurement and those of the various defect classes, using the Euclidean distance as we did in the clustering algorithm. The cluster to which the test measurement belongs should yield the smallest distance value. An extra check is performed by calculating the correlation between the test signal and the various class signals. This time the correct class must result in the highest correlation coefficient. The signals we used for these tests are shown in Figure 20 and the resulting distances and correlation coefficients are summarized in Tables 3 and 4. By comparing these values with our expectations based on visual inspection of the test signals, we conclude that the classification process performs well.

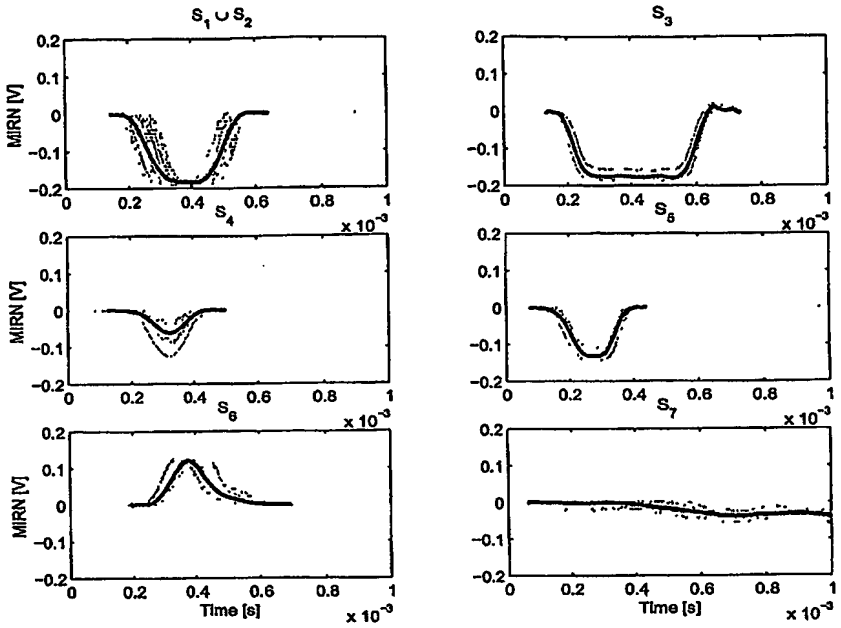


Figure 19: Multivariate 15^{th} order polynomial fit for clustered disc defect signals.

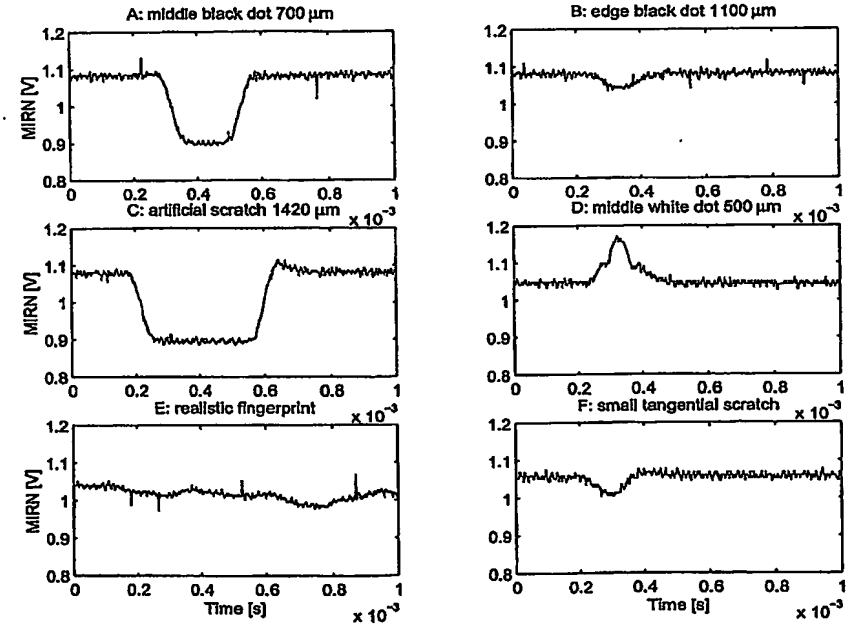


Figure 20: Various test measurements used for class validation.

Table 3: Euclidean distance between property vectors of class signals and test measurements A, ..., F from Figure 20.

Class	Signal A	B	C	D	E	F
$\overline{\mathcal{S}_1 \cup \mathcal{S}_2}$	**370**	906	373	1280	1214	916
$\overline{\mathcal{S}_3}$	604	1250	**214**	1592	1164	1259
$\overline{\mathcal{S}_4}$	839	**258**	1108	548	1363	**244**
$\overline{\mathcal{S}_5}$	503	439	842	877	1410	452
$\overline{\mathcal{S}_6}$	1400	759	1509	**445**	1195	731
$\overline{\mathcal{S}_7}$	1721	1732	1388	1689	**179**	1722

Table 4: Correlation between class signals and test measurements.

Class	Signal A	B	C	D	E	F
$\overline{\mathcal{S}_1 \cup \mathcal{S}_2}$	**0,983**	0,850	0,854	0,274	0,332	0,738
$\overline{\mathcal{S}_3}$	0,768	0,605	**0,999**	0,364	0,444	0,538
$\overline{\mathcal{S}_4}$	0,896	**0,998**	0,603	0,192	0,202	**0,968**
$\overline{\mathcal{S}_5}$	0,945	0,983	0,654	0,208	0,277	0,934
$\overline{\mathcal{S}_6}$	0,279	0,220	0,387	**0,985**	0,388	0,304
$\overline{\mathcal{S}_7}$	0,627	0,521	0,863	0,494	**0,815**	0,408

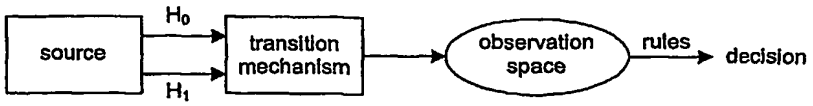


Figure 21: General structure of the detection problem.

4 Defect detection

Various control strategies are available in order to improve playability with respect to disc defects. The succes of all these techniques depends on the ability to detect those specific disturbances in time to take the required countermeasures. When information on the type of defect is available it further becomes possible to select the most suitable strategy. This detection and, closely related, identification of disc defects are the subjects of this chapter.

4.1 The detection problem

4.1.1 Concept and general structure

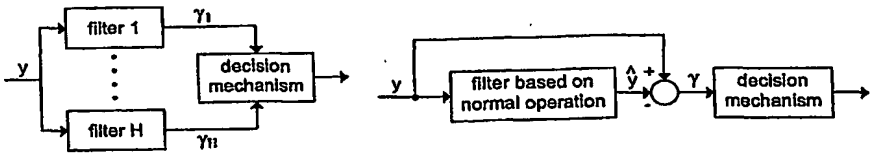
The structure of the detection problem which we will deal with is the following. Given a signal record $(y_1, y_2, \ldots, y_n)$, decide which of the two hypotheses H_0 or H_1 is true:

H_0 : $(y_1, y_2, \ldots, y_n)$ follows the model $S_{\theta,0}$
H_1 : there exists a time instant r, $2 \leq r \leq n$, such that
$(y_1, \ldots, y_{k_d-1})$ follows the model $S_{\theta,0}$
$(y_{k_d}, \ldots, y_n)$ follows the model $S_{\theta,1}$

Here S_θ is a family of models parameterized by the vector θ. See also Figure 21. These models and the signal record together form the 'source' block in Figure 21. The *transition mechanism* maps the hypotheses for a given source into an *observation space*. This mapping follows from the criteria to which the detection must comply. The selection of the valid hypothesis is done by applying a *decision rule* to the mapping result.

Note the subtle difference of the above signal record with the mentioned time series in Chapter 3. When we would choose the record length n equal to that of the whole time series N, the detection structure would prohibit the search for multiple detections between 0 and N. By choosing the size of the observation window $n \leq N$, a sequential search for individual detections can be performed in each part $(y_k, y_{k+1}, \ldots, y_{k+n})$, $k = 0, n+1, 2n+1, \ldots, N$ of a complete time series. Note that in this chapter the symbol i represents an index instead of an element of a set.

For online detection it is important to realize that we are always dealing with a *causal* system. This implies that it is impossible to detect an anomaly precisely at the moment that it occurs. Some delay Δt is inherently present between the detection at $t = k_d$ and the actual occurrence at $t = k_d - \Delta t$ of the anomaly. The goal of a detection system now is to detect a change as quickly as possible after it has occurred, in order that, at each time instant, at most one change has to be detected between the previous detection and the current time point n.



(a) Multiple filter structure

(b) Residual filter structure

Figure 22: General structures for the detection of signal changes.

Finally we remark the similarity between the detection problem and classification or identification. In both cases the behavior of a dynamic system, represented by a signal, is compared with known types of behavior. Based on this comparison a decision is made according to some rules. For detection the decision is whether there is an anomaly present or not while in the case of classification or identification we decide to which class the behavior (signal) belongs.

The generalized form of the above detection structure is the so-called multiple filter structure as depicted in Figure 22a. The observations $y(k)$ with $k = 1, 2, \ldots, N$, are processed by a bank of filters, each of which is based on a particular hypothesis. For instance in the mentioned structure Filter 1, associated with H_0, assumes no change has occurred and Filter 2, which is associated with H_1 assumes that a particular type of change has occurred at a certain time instant. The outputs of the filters, γ, represent signals that should typically be small if the corresponding hypotheses are in fact correct. The decision mechanism is therefore in essence based on determining which of the filters is doing the 'best' job of keeping the corresponding γ's small.

Another general structure for the detection of abrupt changes is the residual-based structure, also illustrated in Figure 22b. In this case a filter is designed based on the assumption that no abrupt change has occurred or will occur. The filter produces a prediction $\hat{y}$ of the observed signal y, based on this assumption and the history of the observed signal. This prediction is subtracted from the actual signal to produce a residual signal γ. If no abrupt change has occurred, γ should be small. Consequently deviations from this behavior are indicative for anomalies, and it is on this fact that the decision mechanism is based.

4.1.2 Defect detection and identification

In the previous section we mentioned three cases in which the need for detection of abrupt changes arise. For disc defect detection in optical disc drives both the second and third reason apply. Since our goal is to prevent that erroneous information will influence the operation of the system, disc defect detection can be seen as an alarm that initializes appropriate countermeasures. Known countermeasures all adjust or completely replace the servo controller in one way or the other, effectively adjusting the tracking mechanism of the read out unit. This is in line with the third reason we mentioned.

From the goal we stated, various requirements can be derived which a disc

defect detection mechanism must meet. In the first place the detection time should be as short as possible. Second we require the highest possible accuracy or reliability. It will be clear that those requirements are by nature conflicting. The search for an optimal detection algorithm with respect to speed and accuracy is further complicated by the robustness demands. Despite the vast amount of different disc defects and the lack of information on other influencing parameters (disc reflection, temperature, substrate thickness e.g.) the detection algorithm must satisfy the first two requirements.

Next to timely detection of a disc defect, we also want to know the type of disc defect. With the general multiple filter structure from Figure 22a both the detection and identification of disc defects can be combined. This can be achieved by formulating a hypothesis for each filter that assumes the presence of one of the defect types that we wish to distinguish. The same can be achieved with the residual filter structure when several of these filters are used in parallel. The resulting defect classes $\mathcal{S}_c$, $c \in \mathcal{C}$ of a classification like we discussed in Chapter 3, can form a good starting point in formulating the required hypotheses.

Decoupling of the detection and identification of disc defects is also possible. This implies that we need one algorithm that is able to detect all different types of disc defects. This relaxes the need for fast defect *identification* and hence its accuracy can be improved. However the chance of false alarms during the defect *detection* increases. Since the detector must be able to detect *all* defects its resolution will be reduced. This makes it harder to distinguish disc defects from other signal distortions. However when the countermeasures initiated by a defect detection do not endanger the proper functioning of the drive in case of a false alarm, the decreased reliability of the detector becomes of less importance.

4.2 Disc defect detection method

Independent from the type of detection method and structure, the goal of disc defect detection implies the need for an on-line algorithm. Off-line detection is not feasible due to the stringent demands on startup times for optical disc drives. There is simply no time to scan the whole disc and memorize the location of all disc defects 'off-line' before the actual data read out starts. However for the selection of affected signal regions as mentioned in Chapter 3, an off-line detection algorithm would suffice.

Various on-line detection methods that fit one of the general forms of Figure 22 have been developed. In this section we present a disc defect detection method that is based on the well-known concept of maximum likelihood (ML). The corresponding theory is treated extensively in literature and already work is done on implementing detectors of this kind in radar and hard-disk drive applications. The method leads to a detector that is easy implementable on-line and that is closely related to defect classification.

4.2.1 Maximum likelihood detection

The method we will discuss here uses the MIRN signal as input. The task of the detector will be to detect whether the influence of a defect, represented by the set of defect classes $\mathcal{S}_\mathcal{C}$, is present in this signal. For the sake of clarity we only consider two possibilities. The first is that no defect is present and the second is

that a defect, represented by $\overline{S}_c$, is present. As discussed in the previous section the detector can easily be extended to cover all defect types by using several detectors in parallel.

We start by defining the two corresponding hypotheses H_0 and H_1 for the disc defect detection problem. The null hypothesis H_0 states that no disc defect is present and H_1 is true when a disc defect is present. The observations of the MIRN signal under the two hypotheses are:

$$H_0 : y(t_s + k) = y_n(t_s + k) \quad (16)$$

$$H_1 : y(t_s + k) = y_n(t_s + k) + y_c(k) \quad (17)$$

with $i = 1, 2, \ldots, N$ determining the detection window $\mathbf{y}(t_s) = (y(t_s + 1), y(t_s + 2), \ldots, y(t_s + N))$. The MIRN signal is modelled as a stochastic process $y(t) = \mu(t) + v(t)$ with a purely deterministic part $\mu(t)$ and a stochastic part $v(t)$. The signal $y_c(k)$ denotes the defect signal obtained from $\overline{S}_c$, and t_s is the defect arrival time. The observations of the MIRN signal and the defect signal are jointly represented as the *source* in Figure 21. Attached to the two hypotheses are the two conditional probability densities $p_{\mathbf{y}(t_s)|H_0}(\mathbf{y}|H_0)$ and $p_{\mathbf{y}(t_s)|H_1}(\mathbf{y}|H_1)$. They define the chance on respectively H_0 and H_1, given the actual observations of $\mathbf{y}_s$.

In order to determine which of the two hypotheses is true a *decision rule* is needed. The requirement for such a rule is that it maximizes the reliability of the decision for a given detection time. Stated differently it must minimize the detection time for a given level of reliability. We now assume that the chance of a false alarm[2] and that of a missed detection[3] are directly related to the detection time or, with a given sample time, the size of the detection window. In that situation the *likelihood ratio test* yields an optimal decision rule with respect to those criteria. It is defined as:

$$\frac{p_{\mathbf{y}(t_s)|H_1}(\mathbf{y}|H_1)}{p_{\mathbf{y}(t_s)|H_0}(\mathbf{y}|H_0)} \; >^{H_1} \leq^{H_0} \; \eta \quad (18)$$

where H_1 is accepted when the ratio in the left-hand side of (18) is greater than the threshold η. Else H_1 is rejected, indicating that no defect is detected. The likelihood ratio forms the probabilistic *transition mechanism* while the threshold comparison is the *decision rule* according to Figure 21.

For simplicity we now assume that the normal, unaffected MIRN signal is an uncorrelated, zero-mean stochastic process (Gaussian white noise) with variance λ. In that case the likelihood ratio test for the presence of a disc defect is:

$$\sum_{k=1}^{N} y(t_s + k) y_c(k) \; >^{H_1} <^{H_0} \; \frac{1}{2} \sum_{k=1}^{N} y_c^2(k) + \lambda \cdot \ln \eta \quad (19)$$

which can be written in the form of a simple discrete time FIR-filter. The detector then becomes:

$$\sum_{k=0}^{N-1} Y_c(N - k) \left(z^{-k} \cdot Y(z) \right) \; >^{H_1} <^{H_0} \; TH \quad (20)$$

[2]Rejecting H_0 when it is true; type I error

[3]Accepting H_0 when it is false; type II error

where TH denotes a new threshold value.

The assumption that the unaffected MIRN signal can be described by Gaussian white noise is not a very realistic one. A more realistic representation can be obtained by incorporating the coloring of the noise for the quasi-stationary MIRN signal. The required changes in the FIR-filter of (20) are specified in literature. However, for reasons of simplicity, we continue to use the white noise assumption in the remainder of this chapter.

The choice of the threshold value TH and the detection window size N depend on the requirements of detection speed and reliability. These requirements on their turn depend on other elements of the optical disc drive such as the used control strategy during disc defects, the data decoding and error correction algorithms. More research and experimental validation are needed on the integrated system in order to determine suitable values for these important parameters.

4.2.2 Some intuitive adjustments

The FIR-filter from (20) forms the core of the maximum likelihood detector. Basically the output of this filter is a multiplication of N samples of the input signal with N corresponding samples of the assumed defect reference signal. We already suggested to use the time series $\overline{\mathcal{S}}_c$ as reference signals $y_c(k)$ in c parallel detection filters. In Chapter 3 we mentioned that the edges of these time series, obtained from fitted polynomials, are not very accurate. Another disadvantage of these signals is that they hardly show any deviations from the normal MIRN signal during the first few samples. When using these signals in the suggested detector this will result, when N is chosen small as required by the demands for fast detection, in an output that will hardly show any change when a defect is present. Therefore we decide to approximate the edges of the fitted defect signal with a straight line as depicted in Figure 23 for one of the reference signals.

All the models are adjusted like we discussed and simulations are done with the resulting FIR-filter. For one of these simulations the resulting output of the filter is shown in Figure 24. We remark that this output is approximately 20 times higher than for the initially used defect class models with the smooth edges (gray line in Figure 23). The actual detection takes place by searching for the time instant where one of the output signals reaches the threshold level TH. The figure further shows that for all defect models $y_c(k)$ the filter gives a significant output. From this observation we conclude that the distinctive capabilities of the various filters (one for each defect class as identified in Chapter 3) is low. Further investigations show that, independent of the class to which a test signal belongs, the reference signal for $\mathcal{S}_1 \cup \mathcal{S}_2$ results in the highest filter output and that for $\mathcal{S}_7$ in the lowest.

From the above we conclude that a combined defect detection and identification approach does not give satisfactory results. More accurate models, especially at the starting edge of the affected signals and better knowledge of the noise coloring and probability distributions could possibly improve the performance. However we now chose another approach. We observe that the slope of the reference signal $y_c(k)$ in the detection window is the feature that determines the amplitude of the FIR-filter response. In the first N samples it appears that the reference signal for the defect class $\mathcal{S}_1 \cup \mathcal{S}_2$ has the steepest slope of all signals. See also Figure 24 where the output of the FIR-filter is shown for the

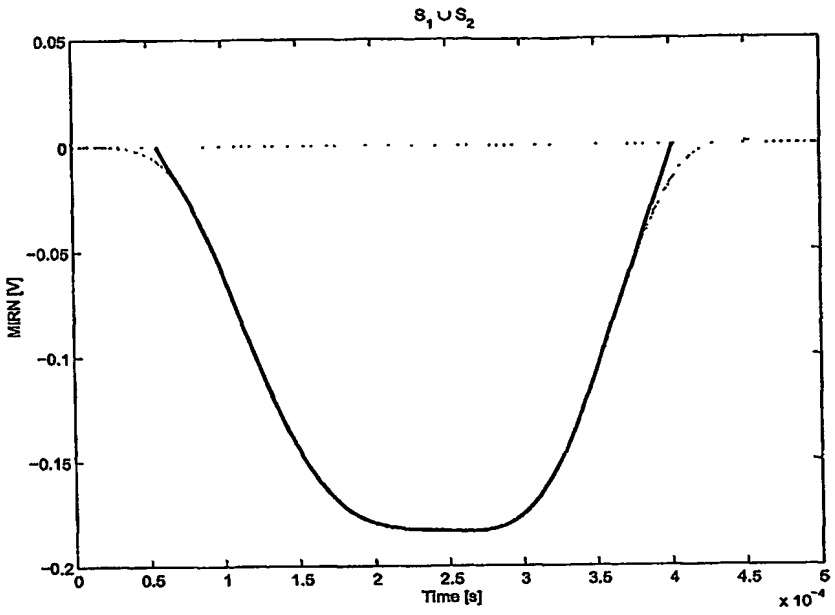


Figure 23: Straight line approximation for the reference signal edges.

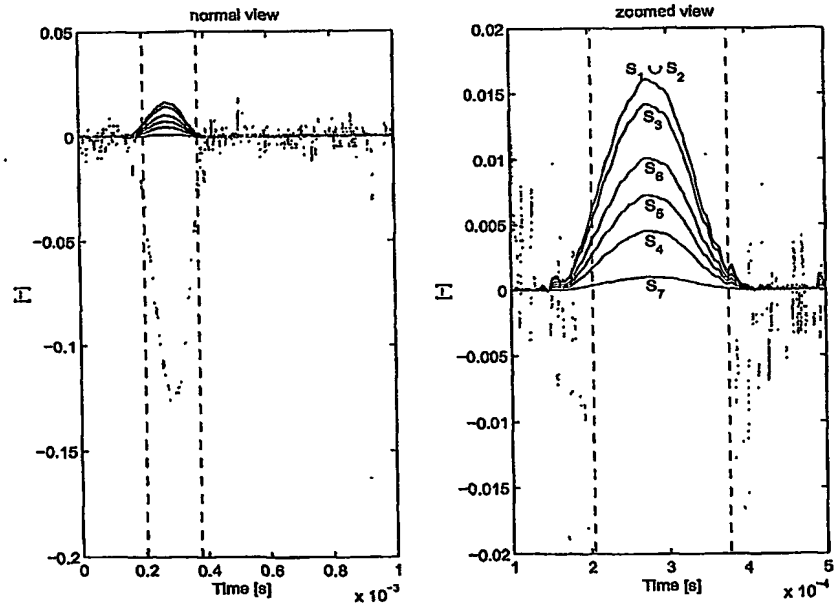


Figure 24: FIR-filter output for defect class reference signals; 700 μm quarter black dot, $N = 10$.

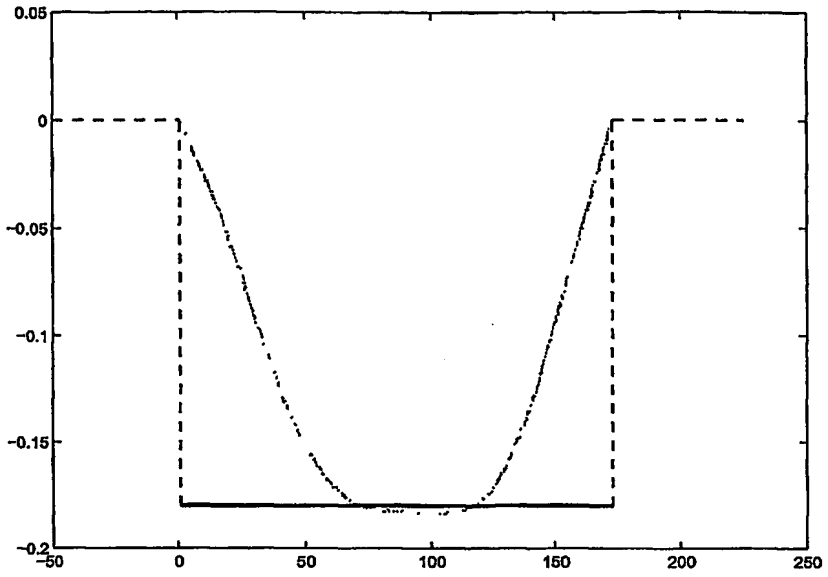


Figure 25: Reference signal with infinite slope for defect detector.

different reference signals.

The idea now is to model the defect with a signal that has an infinitely steep slope. The amplitude of this 'block form' defect model can be chosen at will, as long as the corresponding threshold value is adjusted accordingly. The new defect model is depicted in Figure 25. Simulation results with this new reference signal are shown in Figure 26. Here we clearly show the improvement in detection speed that can be achieved by using this abstract defect model. Note that when the amplitude of the block form defect signal is chosen equal to one, the output of the FIR-filter is reduced to a simple summation of N samples of the incoming MIRN signal.

4.2.3 Disc defect identification method

The choice for the 'infinite slope' reference signal also implies that an extra algorithm is needed to identify the exact type of disc defect whenever one is detected. A logical choice is to use the MIRN signal mapping we developed for the defect clustering algorithm. As soon as a defect is detected by the defect filter, we start to construct the property vector p for the incoming MIRN signal. See also Section 3.2.2 and 3.2.3. Initially only N samples of the signal are available but for each new sample extra information becomes available and hence the estimates of the various properties become more accurate. For the reference signals $\overline{S}_c$ this procedure can be performed off-line, resulting in c property matrices or look-up tables, denoted by P_c. Each row n, $n = 1, 2, 3, \ldots$ of such a matrix holds the property vector for the first $N + n - 1$ samples of the corresponding reference signal. At each time instant k we now can calculate the Euclidean distance between the property vector p of the input signal and those of all the

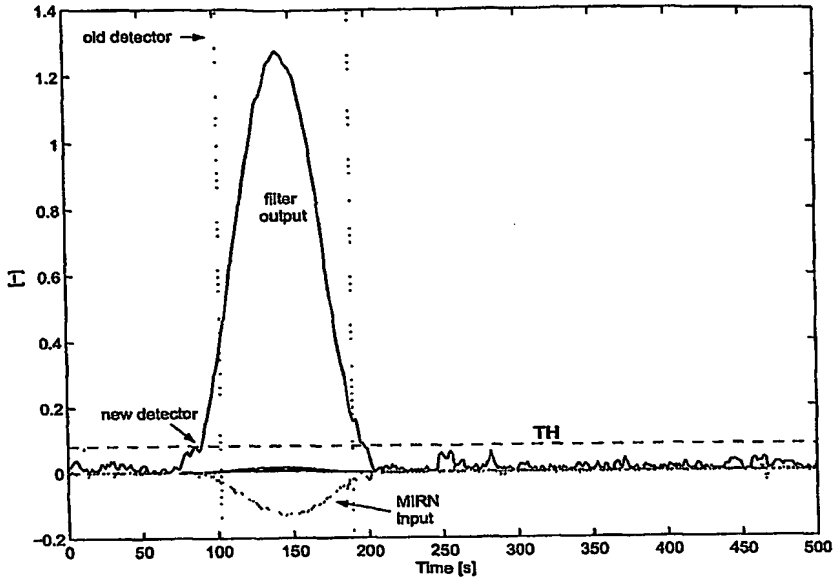


Figure 26: FIR-filter output for block form reference signal; 700 μm quarter black dot, $N = 10$.

class reference signals $P_{c,n}$, $c \in \mathcal{C}$. When the number of available samples is sufficiently high, one of these distances will become significantly smaller, thereby identifying the occurring disc defect on-line.

4.3 Detection performance

In this section we present a simulation model of the defect detection and identification algorithm that we discussed in the previous section. With this model we try to assess whether the intuitive design choices we made, result in an improved defect detector. Next to the validation of the method we also discuss various practical issues that are of importance for the integration of the algorithm in optical disc drives.

4.3.1 Simulation results

The simulation model of the defect detection and identification algorithm is made in Simulink. As much as possible we restrict ourselves to the use of elementary blocks like summation points, switches, comparators and unit delay blocks. This approach results in a basic model for which translation to a real hardware implementation is relatively easy. Details about the simulation model can be found in Appendix 6.

In Figure 27 the simulation results are shown for a quarter black dot of 900 μm. The graph shows that the new detector responds about 30 μs faster to the defect than the currently implemented detector. This improvement in detection speed goes at the cost of some 'false' alarms. As discussed in Sec-

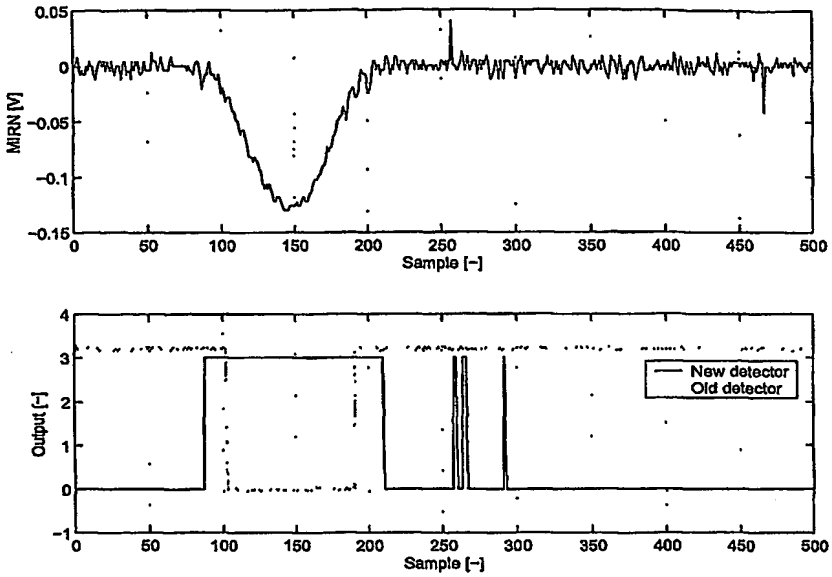


Figure 27: ML detector simulation result for a 900 μm quarter black dot; $N = 10$, $TH = 0.5$.

tion 4.2.1 an optimal threshold value can be determined that gives the best trade-off between detection time and false alarms.

Experiments with other values for the detection window size N revealed no significant changes in the resulting outputs of the simulation model. The corresponding result of the implemented identification algorithm for the black dot simulation is shown in Figure 28. From this graph it becomes clear that in approximately 45 μs the correct defect type—the class to which the defect belongs—can be determined.

4.3.2 Validation

Simulation results already showed that the maximum likelihood detector with the block form reference signal performs equally good or better than the currently used detector. However we also want to compare the results of the detector with the simple case in which defects are detected whenever the MIRN signal itself passes a certain threshold level. In order to do this we adjust the threshold levels for both methods so that no false alarms will occur during the simulation. With these TH values we then compare the resulting detection times for the ML detector and the direct method. We perform simulations for several different test defects, where for each defect three different MIRN measurements are used. The results of these comparisons are summarized in Table 5. From this table we can conclude that the ML detector gives an output signal with a better signal-to-noise ratio. Therefore the threshold level can be set lower in this case, yielding faster detection. From the experiments we also conclude that in all cases the identification algorithm indicates the correct defect class for the

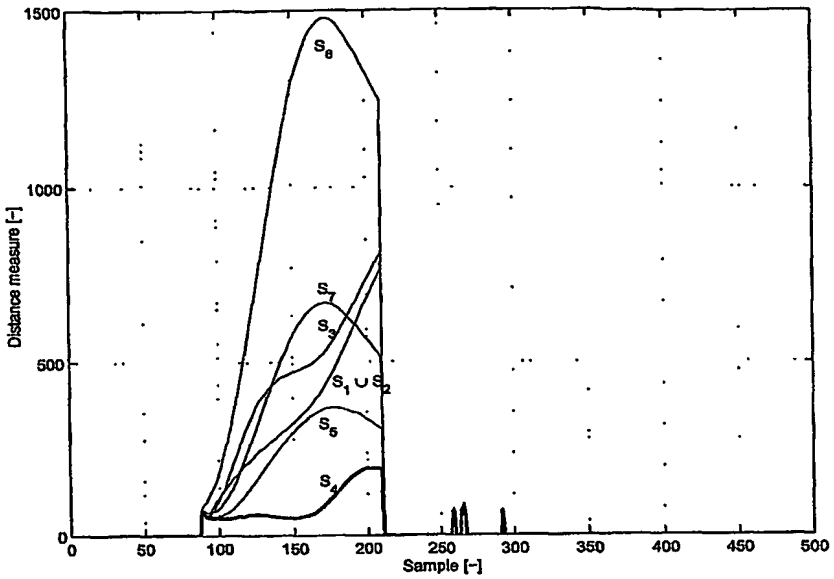


Figure 28: Defect identifier simulation result for a 900 μm quarter black dot; N = 10, TH = 0.5, Euclidean distance measure.

test measurements, within half of the total defect duration time.

4.3.3 Implementation

The simulation model is constructed in such a way that it can form the basis for the actual implementation of the algorithm. The most important issue that requires attention during implementation is the offset cancellation of the MIRN signal. During simulations we use MIRN signals from which the DC offset is removed. See also Section 3.2.1. Tests with the simulation model show that the method is very sensitive to these offsets in the MIRN signal. As discussed in Section 3.2.1 the offset can be determined by calculating the average value of the MIRN signal when it is unaffected by any disturbances. In an on-line implementation the required mean value can be calculated from a fixed number of unaffected samples and it can be updated repetitively. Furthermore a good initial offset value must be available that, for instance, is determined during the drive's initialization sequence.

Another issue that must be taken into consideration when implementing a detector in a (re-)writable drive is the laser power adjustment. When an optical disc drive switches from write mode to read mode or vice versa, the laser is switched between high and low power. This adjustment causes a severe change in the MIRN signal level to which a defect detector, incorrectly, will react. An easy way to deal with this phenomenon is to ignore the defect detector output for a short period of time whenever a laser power adjustment takes place.

Table 5: Detection time difference for ML and direct threshold methods; (for negative values ML detection is faster).

Test defect	Experiment 1	Experiment 2	Experiment 3
edge black dot 700 μm	-24 μs	-20 μs	-24 μs
quarter black dot 900 μm	-30 μs	-30 μs	-10 μs
middle black dot 1100 μm	0 μs	-14 μs	-6 μs
scratch 520 μm	-4 μs	-2 μs	-14 μs
scratch 1020 μm	-3 μs	4 μs	-2 μs
scratch 1520 μm	-2 μs	0 μs	0 μs
scratch at $R = 32mm$	-62 μs	-64 μs	—[1]
scratch at $R = 32mm$	0 μs	-16 μs	-2 μs
scratch at $R = 35mm$	-12 μs	-4 μs	-16 μs
edge white dot 1000 μm	-12 μs	-16 μs	-14 μs
middle white dot 1000 μm	-28 μs	-28 μs	-10 μs
normal fingerprint	-134 μs	-50 μs	-40 μs
heavy fingerprint	0 μs	-10 μs	-24 μs
realistic fingerprint	-38 μs	0 μs	-32 μs

[1] Time difference unknown since no detection occurred for the direct threshold method with the TH value, required to prevent false alarms.

5 Conclusions

With respect to disc defect classification the following can be concluded:

- The hierarchy is a suitable disc defect classification structure as discussed in Section 3.1 and 3.3.
- With a combination of signal theory and physical insight, an usable mapping of defect measurements into a property space is obtained, which is presented in Section 3.2.2.
- As shown in Section 3.3, the Euclidean distance and Ward linkage are geometric dissimilarity measures that are usable in distinguishing different types of disc defects.
- The developed classification method did not give satisfactory results when applied to other servo signals as became clear in Section 3.3.
- The fitted polynomials that were used to describe the defect classes, presented in Section 3.4.1, are far from perfect but usable in classifying new disc defects as followed from Section 3.4.2.

The following can be concluded about disc defect detection in general and the presented approach in particular:

- Detection is a form of classification or identification as became clear in Section 4.1.
- Causality in on-line detection methods implies a trade-off between detection speed and reliability as was shown in Section 4.1.1.
- The method of maximum likelihood presented in Section 4.2.1, provides a manner to obtain an optimal performing detection algorithm with respect to speed and reliability.
- The reference signals describing the obtained defect classes are not suitable for usage in a detection structure with multiple ML detectors, which followed from Section 4.2.2.
- Disc defect detection was improved by using a single ML detector that is based on a 'block form' defect reference model like we discussed in Section 4.3.1 and 4.3.2.
- The investigated detection method is very sensitive to signal offsets in the MIRN signal as mentioned in Section 4.3.3.
- On-line disc defect identification based on the defect clustering algorithm from Chapter 3 seemed accurately and fast enough for real time implementation as discussed in Section 4.2.3 and 4.3.

6 Detector simulation model

Here we present the most important elements of the Simulink model that is used to simulate the behavior of the maximum likelihood defect detector.

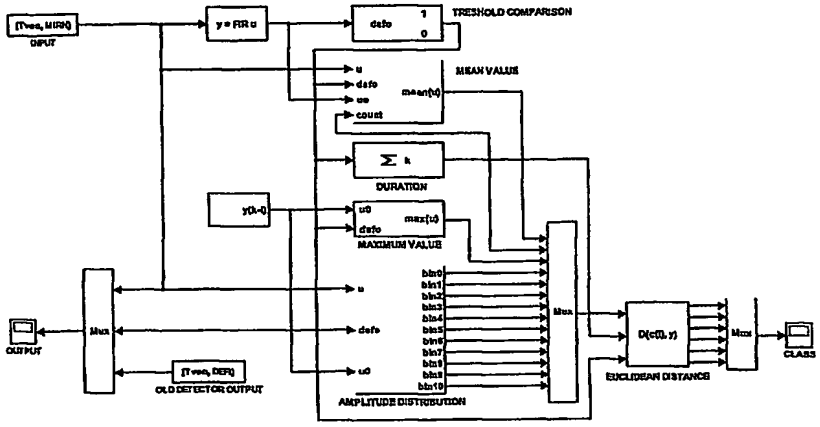


Figure 29: Root level diagram of the ML defect detector simulation model.

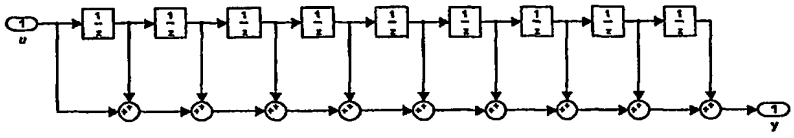


Figure 30: FIR-filter of the ML defect detector simulation model.

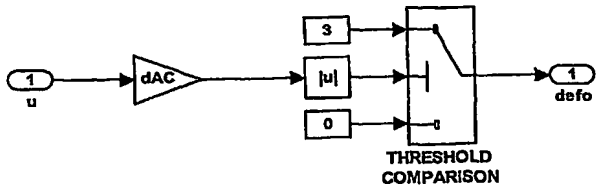


Figure 31: Threshold comparator of the ML defect detector simulation model.

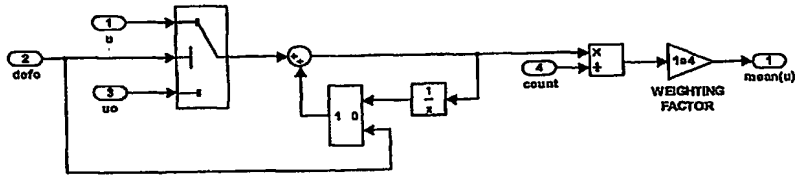


Figure 32: Mean value calculation of the ML defect detector simulation model.

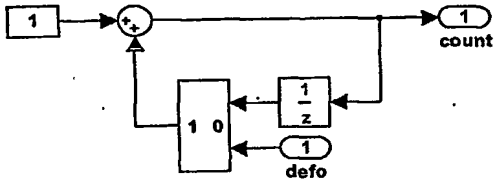


Figure 33: Sample counter of the ML defect detector simulation model.

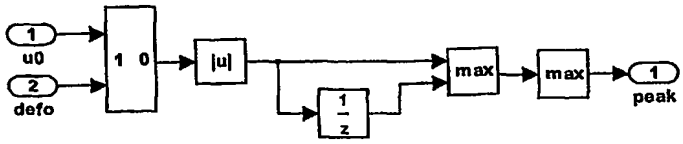


Figure 34: Peak value calculation of the ML defect detector simulation model.

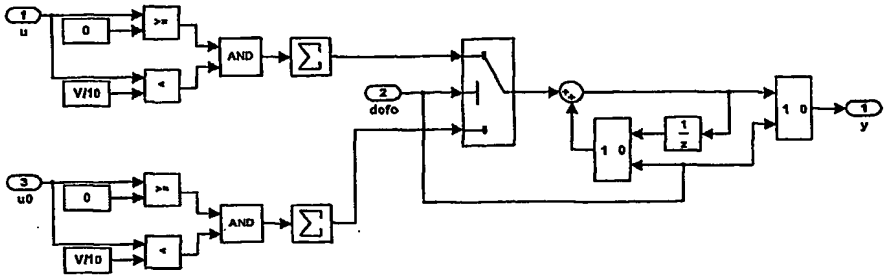


Figure 35: Amplitude distribution calculation of the ML defect detector simulation model.

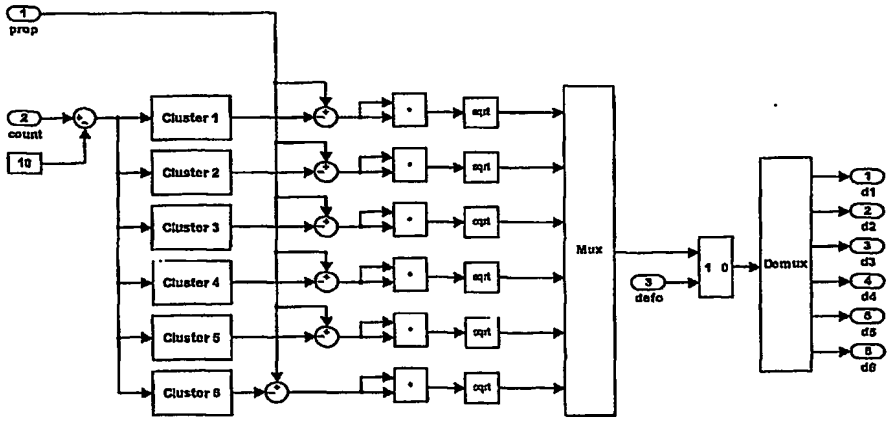


Figure 36: Euclidean distance calculation of the ML defect detector simulation model.

CLAIMS:

1. Device comprising:
means for detecting anomalies in a signal by processing a plurality of samples of the signal and by comparing the processed samples with a threshold.

2. Device as claimed in claim 1, further comprising means for identifying a particular anomaly among a plurality of predetermined anomalies by matching, comparing, or the like, said signal with a plurality of reference signals corresponding to said plurality of predetermined anomalies.

3. Method for determining anomalies in a signal, said method comprising:
processing a plurality of samples of the signal; and
comparing the processed samples with a threshold.

4. Device, system or method substantially as hereinbefore described and/or shown in the figures included therein.